



Lidil

Revue de linguistique et de didactique des langues

53 | 2016

Phraséologie et genres de discours

Repérer les clichés dans les romans sentimentaux grâce à la méthode des « motifs »

Dominique Legallois, Thierry Charnois et Thierry Poibeau



Édition électronique

URL : <http://journals.openedition.org/lidil/3950>

DOI : 10.4000/lidil.3950

ISSN : 1960-6052

Éditeur

UGA Éditions/Université Grenoble Alpes

Édition imprimée

Date de publication : 30 mai 2016

Pagination : 95-117

ISBN : 978-2-84310-326-1

ISSN : 1146-6480

Référence électronique

Dominique Legallois, Thierry Charnois et Thierry Poibeau, « Repérer les clichés dans les romans sentimentaux grâce à la méthode des « motifs » », *Lidil* [En ligne], 53 | 2016, mis en ligne le 01 janvier 2017, consulté le 30 octobre 2020. URL : <http://journals.openedition.org/lidil/3950> ; DOI : <https://doi.org/10.4000/lidil.3950>

© Lidil

Repérer les clichés dans les romans sentimentaux grâce à la méthode des « motifs »

*Dominique Legallois**, *Thierry Charnois***
*et Thierry Poibeau****

RÉSUMÉ

Cet article porte sur la détection automatique de configurations lexico-grammaticales — appelées « motifs » — caractéristiques d'un genre textuel. On s'intéresse plus particulièrement aux motifs exprimant une des dimensions phraséologiques des romans sentimentaux, à savoir le cliché. Deux méthodes de calcul des motifs sont présentées : l'information mutuelle et le calcul des spécificités. L'article donne de nombreux exemples à partir de l'analyse d'un corpus composé de 150 romans contemporains : 50 romans policiers, 50 romans dits « sérieux » et 50 romans sentimentaux. Les résultats obtenus permettent de valider l'idée que les clichés sont exprimés non pas seulement par des séquences figées, mais aussi par des séquences schématiques.

ABSTRACT

This paper deals with the exploration of genre-specific phraseology, more precisely, of “sequential patterns” expressing clichés. In our approach, a sequential pattern (or “motif”) is a specific and regular lexical and grammatical configuration. This unit is automatically detected; genre-specific motifs are calculated on the basis of two alternative methods (mutual information and calcul of specificities). The analysis is based on a large number of illustrations taken from a corpus of 150 contemporary novels: 50 “serious” novels, 50 crime novels and 50 sentimental novels. The study supports our initial hypothesis: sentimental novels are characterized by a high level of clichés, and these clichés are expressed through specific sequential patterns.

* CLESTHIA, Université Sorbonne Nouvelle—Paris 3.

** LIPN, CNRS, Université Paris 13.

*** LATTICE, CNRS, ENS et Université Sorbonne Nouvelle—Paris 3.

Plus que tout autre sous-genre romanesque, le roman sentimental est intrinsèquement lié au cliché. Peut-être est-ce même là sa seule et véritable nature : le roman sentimental s'identifie au cliché, il est lui-même cliché. Pour cette raison, on ne saurait accuser ce type de roman de faillir à des exigences esthétiques ou stylistiques — un tel jugement critique témoignerait d'une méconnaissance du genre. Au contraire, un texte sentimental délesté de formules « toutes faites », en faveur de tournures originales, ne serait tout simplement plus un texte sentimental. Le lyrisme, le romantisme et l'absolu se jouent nécessairement dans une certaine excessivité des formes, tant au niveau sémantique (l'hyperbolisme), qu'au niveau de leur fréquence, leur répétition. Cette excessivité est l'essence même du cliché, dans ce paradoxe quelque peu troublant où la répétition de l'exagéré conduit à la banalité, au conformisme qui nourrit le besoin de déjà-lu du lecteur (de la lectrice ?).

La définition que nous donnons ici du terme *cliché* est relativement classique : il s'agit d'un stéréotype d'expression devenu banal sous l'effet de la répétition. Très souvent, s'ajoute le fait que le cliché est un fait de style :

Pour que « ma flamme » ou « pure comme un lys » rejoignent la catégorie des clichés, il faut qu'ils soient perçus comme des expressions, des effets de style ressassés dans un contexte culturel où ce type de récurrence appelle un jugement dépréciatif. (Amossy & Rosen, 1982, p. 9)

Si les recherches en linguistique et en analyse du discours sur les clichés et les formules stéréotypées font référence presque systématiquement à la paralittérature et au roman *fleur bleue* (Eco, 1979 ; Dufays, 1994 ; Amossy & Herschberg Pierrot, 1997), si les recherches littéraires sur le roman populaire ont bien cerné les caractéristiques historiques, sociologiques et esthétiques du genre sentimental (Péquignot, 1991 ; Houel, 1997 ; Constans, 1999), il reste que l'inventaire des formes invoquées n'est que très partiel, et se limite à quelques locutions ou collocations¹, donc à des unités relativement figées. Par exemple, dans un roman *Harlequin*, Richaudeau (1986) recense :

« Le cœur (de l'héroïne) bondit dans sa poitrine » ; « (ses) paumes moites, son estomac en révolution » ; « l'aéroport (qui) grouillait de passagers », tout cela dans la seule première page du roman. Puis, je prélève « une

1. L'article fondamental de Riffaterre (1964) mentionne principalement de telles formes.

haute silhouette» ; « un hurlement “démoniaque” » ; « courageusement lutte » ; « le sol de marbre brillant comme un miroir » ; « la sonnerie stridente du téléphone » ; « les “petits seins haut placés” » ; « les couverts d’argent (qui) étincelaient » ; « les chromes étincelants » ; « un long regard » ; « un pas souple » ; « un soupir de ravissement ». J’arrête ici ces citations, arrivé à moins du tiers de l’œuvre. (p. 13)

Si le cliché est le plus souvent considéré pour sa dimension figée — le cliché comme expression toute faite —, nous voudrions ici montrer qu’on peut également aborder la question sous l’angle de la flexibilité : un cliché peut être appréhendé à un niveau général, à partir de ses diverses réalisations, qui toutes ont un air de famille, partageant par exemple un patron syntaxique et des classes lexicales communs. Par exemple, pour des groupes nominaux tels que *la pression de son genou contre le sien / le contact de ses lèvres contre les siennes*, on préférera voir, plutôt que deux clichés, un seul cliché avec deux réalisations différentes. Il convient d’identifier, sur corpus de grande taille, le patron lexico-grammatical (noté **le NC de DETPOSS NC contre le sien**²) sous-jacent aux réalisations effectives.

Nous présentons dans cet article³ une méthode qui permet d’extraire automatiquement de telles configurations lexico-grammaticales schématiques, candidates à la qualification de cliché. Cette méthode, fondée sur l’identification de « motifs séquentiels », est encore peu utilisée en linguistique de corpus ou en textométrie ; elle reste, à ce stade de développement, largement perfectible. Néanmoins, nous espérons montrer ici la pertinence de son application pour la caractérisation phraséologique des genres et des styles.

La teneur de l’article est donc principalement méthodologique ; le rapport entre roman sentimental et cliché sera illustré par de nombreuses formes, plutôt qu’il ne sera véritablement étudié dans l’optique d’une caractérisation précise du sous-genre.

La première partie présente le corpus analysé, ainsi que quelques résultats obtenus par extraction des segments répétés. Les segments répétés sont en principe de bons candidats pour le statut d’unités phraséologiques. Cependant, la deuxième partie montre que des unités moins figées, s’apparentant à des patrons lexico-grammaticaux (tel que celui présenté plus haut), que l’on désigne également sous le terme de motifs

2. NC = nom commun, DETPOSS = déterminant possessif.

3. Quelle meilleure revue que *Lidil* pour évoquer le roman sentimental ?

séquentiels, peuvent être automatiquement extraites, soit par calcul « endogène », soit par calcul « relatif ». La troisième partie présente les résultats de l'extraction de motifs, et propose un classement de ces formes.

1. Présentation du corpus et analyse en segments répétés

Le corpus est composé de 150 romans contemporains : 50 romans policiers, 50 romans dits « sérieux », 50 romans sentimentaux. Voici, à titre indicatif, quelques noms d'auteurs de référence pour chacun des sous-genres. Le sous-genre « sérieux » est composé d'œuvres d'auteurs tels que Le Clézio, Rouaud, Djian, Carrère, Modiano, etc. Le terme de *roman sérieux* est utilisé par Malrieu et Rastier (2001). L'étiquette est commode, mais ne fait pas vraiment sens d'un point de vue générique, puisque l'ensemble des textes n'est pas nécessairement homogène. Le sous-genre policier comprend, entre autres, des romans de Jonquet, Izzo, Thery, Vargas, Manchette, etc. Le corpus roman sentimental est uniquement composé de romans Harlequin, éditeur par excellence des romans dits sentimentaux. Ces romans sont des traductions de l'anglais⁴. Policiers et romans sérieux servent de corpus de référence pour établir des contrastes avec le sous-genre sentimental — notre corpus cible.

L'extraction de motifs s'inscrit dans une démarche générale de caractérisation linguistique des textes ; aussi est-elle appropriée pour analyser et décrire les genres de discours, ou bien les styles d'auteurs, quoique son application puisse être étendue à d'autres types d'étude. Surtout, la méthode des motifs est conçue comme complémentaire à celles des analyses en unités morpho-syntaxiques discrètes popularisées par les travaux de Biber (1988)⁵. Comme on le sait, l'analyse discrète s'appuie sur des descripteurs (principalement des catégories grammaticales, des catégories sémantiques, les lemmes, etc.) dont on mesure par statistiques les sur-emplois dans un texte ou ensemble de textes, par rapport à un autre ensemble de textes. Depuis plus de trente ans, les analyses en textométrie ou en linguistique de corpus, fondées sur l'approche « discrète », ont permis de découvrir des spécificités génériques, jusque-là peu visibles pour les analyses non informatisées et non

4. Un travail nécessaire serait d'évaluer l'incidence de la traduction sur la phraséologie.

5. Déjà Sueur en 1982 proposait ce type d'analyse en France.

statistiques. Des corrélations entre les descripteurs ont pu être calculées, qui donnent une représentation de la façon dont certaines formes sont complémentaires les unes des autres.

Néanmoins, le besoin s'est fait sentir d'apprécier également des unités, non plus discrètes, mais syntagmatiques ; Salem (1987) propose une analyse en segments répétés (SR), qui prend en compte les suites récurrentes de n-unités lexicales — c'est ce type d'analyse que nous appliquons ici.

Par exemple, le tableau suivant recense les 39 SR de 5 et 6 formes, avec une fréquence égale ou supérieure à 10 et apparaissant dans au moins 25 romans sentimentaux (les textes ne sont pas lemmatisés).

que se passe t il	149	pour la première fois de	51
la prit dans ses bras	100	ce n'est pas ce que	51
dans la salle de bains	97	je ne sais pas comment	50
sur la pointe des pieds	91	il jeta un coup d'œil	50
leva les yeux au ciel	85	lui prit la main et	50
avec un peu de chance	76	n'a rien à voir avec	49
elle ne put s'empêcher de	72	la porte de la chambre	49
pour la première fois depuis	67	il ne lui restait plus	48
il n'y a pas de	65	il se tourna vers elle	48
et se dirigea vers la	65	il la prit dans ses	47
la prendre dans ses bras	65	je ne sais pas ce	46
se dirigea vers la porte	64	mais je ne peux pas	45
je ne veux pas que	62	la première fois de sa	44
que je ne suis pas	59	je ne pense pas que	43
de l'autre côté de la	59	était sur le point de	43
de la tête aux pieds	56	première fois de sa vie	42
il se pencha vers elle	55	je me suis dit que	42
de la salle de bains	54	je lui ai dit que	39
t elle avec un sourire	52	au dessus de sa tête	31
elle se tourna vers lui	52		

Tableau 1. – Les SR du sous-genre roman sentimental
(apparaissant au moins 10 fois dans au moins 25 romans).

À titre de comparaison, les résultats pour le sous-genre policier sont :

je ne sais pas si	76
je ne sais pas ce	68
de l'autre côté de la	75
pas eu le temps de	50
il n'y a pas de	84
il n'y avait pas de	56

Tableau 2. – Les SR du sous-genre policier
(apparaissant au moins 10 fois dans au moins 25 romans).

Pour le sous-genre roman sérieux, un seul segment est identifié : *de l'autre côté de la* (8 fois).

Comparé aux deux autres sous-genres⁶, le sous-genre roman sentimental est fortement affecté par des segments répétés. On est surpris, non pas tant du score de ce sous-genre, que de la quantité négligeable de segments de longueur 5 ou 6 formes employés dans au moins la moitié des textes policiers et sérieux : ces deux sous-genres ne sont pas marqués par des formes récurrentes, en tous cas de cette dimension. En ce qui concerne le roman sentimental, au contraire, le dispositif des segments répétés montre que les séquences répétitives ont un rôle à jouer dans la caractérisation du sous-genre. Peut-on dire pour autant que les segments sont des clichés ? On remarque bien la présence d'unités telles que *la prit dans ses bras*, *la prendre dans ses bras*, *lui prit la main* et qui ne peuvent constituer, sans un retour au texte, que des indices. Imposer d'autres contraintes à l'extraction peut s'avérer plus fructueux. Dans le tableau 3, figurent des segments répétés de 3-4 formes, calculés par le test de l'information mutuelle (IM : voir plus bas), grâce au logiciel *Collocates* de Barlow (2004). La fréquence brute est donnée entre parenthèses, suivie de la mesure de l'IM. On repère aisément certains clichés. Par exemple : *prit une profonde inspiration*, *poussa un profond soupir*, segments dans lesquels *profond* joue le rôle d'un collocatif exprimant une fonction MAGN (selon la terminologie de la théorie Sens-Texte [Mel'čuk, 1996]). Mais aussi : *se mordit la lèvre*, *lèvres sur les siennes*, *visage entre ses mains*, *en fronçant les sourcils*, *haussa les*

6. Dont nous ne dirons rien de plus, dans cet article, faute de place. Pour une étude du stéréotype dans le roman policier, voir Marion (2009).

épaules qui constituent des expressions récurrentes indiquant des comportements somatiques :

jeta un coup d'œil (256) 27,864832	prit une profonde inspiration (92) 30,055021	poussa un profond soupir (47) 28,200425
quelques minutes plus tard 107 25,696238	une demi-heure plus tard (35) 27,092478	quelques instants plus tard (79) 26,655927
en fronçant les sourcils (44) 24,963201	se mordit la lèvre (67) 25,642335	par-dessus son épaule (110) 25,181271
du bout des doigts (103) 23,366181	du coin de l'œil (50) 23,659888	la pointe des pieds (91) 23,624389
ne put s'empêcher de (194) 22,775916	fronça les sourcils (71) 23,100889	visage entre ses mains (39) 23,031843
oh mon dieu (35) 22,551930	lèvres sur les siennes (50) 22,668404	elle ne put s'empêcher (85) 22,664389
se dirigea vers la porte (64) 22,198303	ce qui s'est passé (115) 22,330827	se laissa tomber sur (58) 22,283598
lui adressa un sourire (104) 21,934598	en se dirigeant vers (50) 22,113307	haussa les épaules (128) 22,006932

Tableau 3. – Les 24 premiers SR (frq min = 10, rép. = 25) calculés par IM dans le sous-genre roman sentimental.

Néanmoins, on peut vouloir s'intéresser à des unités plus schématiques, autrement dit moins lexicalement spécifiées. Ainsi, par exemple, les séquences *prit une profonde inspiration*, *poussa un profond soupir*, peuvent être considérés comme des réalisations différentes d'un même schéma. Aussi la méthode des motifs est-elle proposée pour compenser la dimension trop lexicale des segments répétés.

2. Les motifs

2.1. Principe général

La méthode consiste à identifier des patrons dont la nature des éléments est mixte : des formes fixes, des lemmes et des catégories morpho-syntaxiques (Quiniou, Cellier, Charois & Legallois, 2012 ; Longrée & Mellet, 2013 ; Legallois & Prunet, 2015 ; Diwersy & Legallois, à paraître ; Legallois, à paraître). À l'image des segments répétés, ces patrons sont extraits automatiquement de manière non supervisée, sans que l'analyste propose de requêtes. De façon plus précise, la méthode

consiste à modifier par expressions régulières les sorties d'un catégoriseur tel que *Cordial*, pour en conserver :

- les formes des unités invariantes, telles que les prépositions, les conjonctions, certains adverbes fréquents, etc. ;
- les lemmes de certains verbes très fréquents (auxiliaires, verbes aspectuels, verbes modaux, etc.) et la réduction des pronoms personnels à une forme canonique ;
- les catégories morpho-syntaxiques (nom commun, nom propre, adverbe, verbe, adjectif, etc.). Les tiroirs verbaux sont identifiés.

Le corpus ainsi étiqueté possède une dimension spécifique avec les unités invariantes, mais aussi générique avec les lemmes et plus encore, les catégories. Précisons que les ponctuations sont considérés également comme des descripteurs. Voici un exemple d'étiquetage :

comment pouvoir on être aussi ADJ et ADJ ? pourtant, à le NC de le jour, il devoir bien INF que NP savoir aussi se INF ADJ et ADJ, et surtout très ADJ.

un NC qui le VP ADV, mais que il se VP ADV de INF. quoi que il arriver, trop de NC le VP encore et NP ne être ADV un NC pour lui

Cette annotation correspond au passage suivant :

Comment peut-on être aussi snob et hautaine ? Pourtant, au fil des jours, il doit bien reconnaître que Marnie sait aussi se montrer enthousiaste et généreuse, et surtout très sexy.

Une découverte qui le trouble profondément, mais qu'il s'empresse aussitôt d'ignorer. Quoi qu'il arrive, trop de choses les séparent encore et Marnie ne sera jamais une femme pour lui...

La procédure consiste ensuite à extraire des corpus les segments, de longueur variable. Deux solutions s'offrent alors : soit les segments sont extraits sur la base de l'indice statistique de l'information mutuelle ; soit ils sont extraits simplement avec leur fréquence brute. Dans le premier cas, les segments sont identifiés par calcul endogène, c'est-à-dire sans qu'un genre soit comparé aux autres genres. Dans le second cas, la fréquence des segments de chaque genre est comparée à celle des autres ; le calcul est en somme relatif car fondé sur une comparaison (par calcul des spécificités). Nous considérons ici les deux méthodes comme complémentaires, mais une étude plus précise devra être menée pour évaluer véritablement si l'une est plus pertinente que l'autre.

Quelle que soit la méthode, sont qualifiés de *motifs* les seuls segments présentant une régularité d'ordre lexical (un même paradigme lexical est employé dans le patron — par exemple, les lexèmes relatifs au corps), et/ou fonctionnel : un motif possède une fonction sémantique, pragmatique, rhétorique discernable, voire une fonction d'organisation des plans textuels (comme les trois derniers motifs présentés à la fin de l'article).

Le motif n'est donc pas identifié automatiquement ; il est l'objet d'une interprétation de la part de l'analyste. Par exemple, un segment spécifique des textes sentimentaux tel que **NC, il se V-PS⁷ et** qui se manifeste dans :

- (1) *Repoussant sa **chaise**, **il se leva et** gagna tout de même le couloir pour observer l'échange.*
- (2) *Comme si elle avait senti son **regard**, **elle se retourna et** haussa les épaules avec un sourire triste.*

ne constitue pas une séquence assez homogène (ici sur le plan formel) pour faire unité. Elle n'est donc pas retenue.

2.2. Méthode 1 : extraction des motifs par information mutuelle

L'information mutuelle (IM) est une mesure probabiliste qui favorise les relations rares entre les formes (Church & Hanks, 1990). Les segments partageant un fort score ont tendance à être des groupes figés, assez peu fréquents. Parce que l'étude porte sur le sous-genre et non sur un ou des textes particuliers, nous avons à nouveau appliqué les contraintes d'une répartition minimale des formes dans 25 textes, et d'une fréquence minimale de 10 ; on élimine ainsi les séquences lexicalisées, peu fréquentes, mais que l'IM promeut fortement. Le tableau suivant répertorie les 42 premiers segments de longueur 5 (nous indiquons la fréquence et la mesure de l'IM)⁸ :

7. V-PS = verbe au passé simple.

8. ADJORD = adjectif ordinal ; VIMP = verbe à l'imparfait ; NUM = déterminant numéral ; IMP = impératif ; INF = infinitif ; PRES = participe présent.

pour le ADJORD fois depuis (70) 25,965158	la plupart de le NC (179) 23,145586	avoir rien à voir avec (113) 22,302519
je ne savoir pas comment (59) 21,84964	chaque fois que il VIMP (41) 21,801511	il ne y avoir personne (54) 20,808892
je ne en savoir rien (61) 20,960468	quelque NC plus tard, (165) 20,882687	il ne y avoir aucun (137) 20,68853
ne avoir rien à voir (141) 20,706908	ce être le ADJORD fois (109) 20,704833	ça ne me VP pas (48) 20,41003
ce qui se être passer (216) 20,545691	ne lui rester plus ADV (54) 20,435228	être le ADJORD fois que (94) 20,11894
je ne vouloir pas te (70) 20,171924	je ne savoir pas si (65) 20,120192	NUM NC plus tôt, (57) 19,687173
il ne y avoir rien (143) 19,734185	NUM NC plus tard, (115) 19,731638	je ne arriver pas à (114) 19,508411
ne te IMP pas, (125) 19,62139	le ADJORD fois que je (38) 19,513578	rien à voir avec le (50) 19,488883
il ne lui rester plus (59) 19,49953	pour le ADJORD fois de (57) 19,489862	le ADJORD fois que il (92) 19,195641
tandis que il se VIMP (124) 19,409676	pour le ADJORD fois, (56) 19,397746	ce que je vouloir dire (66) 19,035251
avant de se INF vers (62) 19,144846	je ne penser pas que (84) 19,051509	je ne pouvoir pas te (42) 18,84768
ne vouloir pas te INF (56) 19,025438	il en se PRES vers (74) 18,922056	ne arriver pas à INF (118) 18,733606
ne y avoir aucun NC (135) 18,846788	ne y avoir rien à (39) 18,841015	je ne vouloir pas que (139) 18,533636
NP en se PRES vers (44) 18,712088	je ne pouvoir pas me (51) 18,580391	, ne te IMP pas (53) 18,383526

Tableau 4. – Les 42 premiers segments (frq min = 10, rép. = 25)
calculés par IM dans roman sentimental.

Il arrive qu'une phrase réalise deux motifs, comme dans :

- (3) <Cela n'avait rien à voir avec> <ce qui s'était passé entre eux>
dans la voiture, l'autre jour.

2.3. Méthode 2 : extraction des motifs spécifiques

La deuxième méthode consiste donc à comparer la fréquence des segments patrons lexico-grammaticaux des différents sous-genres en procédant à un calcul des spécificités par la loi hypergéométrique (en utilisant le logiciel R).

Patrons	Frq dans le sous-corpus sentimental	Frq dans les trois sous-corpus
il lui V-PS DETPOSS NC	116	156
il prendre un ADJ NC	101	130
NC, il se sentir	110	160
, V-PS -t-il ADV.	206	234
. ne te IMP pas,	81	109
tandis que il se VIMP	124	166
V-PS le NC pour le	79	105
V-PS NC il. je ne	107	133
, V-PS NP en se PRES	86	111
. il sentir DETPOSS NC	95	120
NC que il lui VIMP	136	183
un NC de NC le V-PS	105	130
V-PS NC il avec un NC	221	246
, répondre NP en PRES	79	103
. sur ce NC,	84	108
il V-PS le NC pour INF	81	105
il lui V-PS le NC.	99	122
il sentir DETPOSS NC se INF	156	166
le NC que il VIMP pour	57	67
il V-PS le NC et se V-PS	111	161
, dire il de un NC ADJ	242	301
il ne pouvoir se INF de INF	126	166
. avec un NC de NC,	80	105

Tableau 5. – Extrait présentant des patrons spécifiques au genre sentimental (tous les segments sont statistiquement spécifiques).

Les patrons spécifiques d'un texte/corpus sont les patrons sur-employés par rapport aux autres textes/corpus. Comme il a déjà été dit, seuls les patrons possédant une régularité en discours (régularité lexicale, fonctionnelle, etc.) sont des motifs. Et parmi ces motifs, tous ne sont pas à considérer comme des clichés. Par exemple, la séquence **PRES le NC, il V-PS** (*Ouvrant la porte, il laissa passer Kara devant lui ; Rouvrant les yeux, elle retint un cri ; Appelant les chiens, il prit le chemin du retour*) ne correspond pas à cliché. Cette séquence peut

cependant être considérée comme motif en raison de sa nature syntaxique (participe présent détaché en position frontale, sujet pronominal et verbe au passé simple).

3. Résultats

Les résultats que nous présentons mélangent les motifs du sous-genre sentimental obtenus par calcul endogène avec IM (ils sont encadrés par < >) et les motifs spécifiques obtenus par comparaison avec les deux autres sous-genres. Les motifs interprétables ne sont pas tous donnés ; nous ne retenons que ceux qui ont un score maximum (noté inf.). Nous indiquons la fréquence brute du motif dans le corpus sentimental, rapportée à la fréquence brute du motif dans les trois corpus. Ne sont commentés que les motifs qui sont, selon nous, révélateurs de clichés.

3.1. *Le motif exprime une relation entre l'actant sujet et l'actant objet*

3.1.1. *La relation est un contact physique*

Une partie du corps de l'actant sujet est en contact avec une partie du corps de l'actant objet :

il V-PS le NC_{corps} sur DETPOSS NC_{corps} (inf. ; 104/126)⁹ :

- (4) *Il referma les mains sur sa taille, et elle l'enfourcha¹⁰.*
- (5) *Il appliqua la bouche sur ses seins, avec douceur, d'abord, puis avec force.*
- (6) *Elle posa les mains sur son torse, et ce simple contact le bouleversa.*

Les parties du corps ont évidemment un rôle fonctionnel important dans l'affaire : il y a les parties investigatrices (*mains, bouche*) et les parties cibles (*taille, seins*). La formule est quelque peu analytique : on

9. « NC_{corps} » est une annotation à postériori. Les NC ne sont pas à priori sémantiquement typés (ils pourraient l'être !), c'est donc dans l'observation des énoncés correspondant à la séquence qu'émerge la particularité sémantique des noms.

10. Et oui ! Harlequin, propose deux collections érotico-pornographiques (Sexy et Spicy : cf. <www.harlequin.fr/collections>) qui, d'ailleurs, ne dénaturent en rien l'aspect sentimental des romans. Seul l'aspect fleur bleue est quelque peu décoloré.

ne prend pas la taille, on n’embrasse pas, on ne touche pas, mais on *referme les mains* ou on *les pose*, et on *applique la bouche* — histoire sans doute de gagner, au final, quelques lignes.

Le contact porte sur une partie du corps de l’actant objet :

il lui V-PS le NC_{corps} (inf. ; 71/85) / **il lui V-PS ADV le NC_{corps}** (inf. ; 50/50)

(7) *Il lui prit les lèvres, l’une après l’autre.*

(8) *Il lui mordilla délicatement la lèvre, la toucha du bout de la langue.*

Ce cliché est directement lié à un stéréotype sexuel : le contact exprimé par ce motif est initié quasi exclusivement par un sujet masculin. Autrement dit, c’est l’homme qui prend l’initiative de la relation charnelle.

L’inchoativité de la relation charnelle est explicite :

<se mettre à lui INF le NC_{corps}> (inf. ; 14/14)

(9) *Quand il se mit à lui caresser les seins, il la sentit retenir son souffle contre ses lèvres.*

(10) *Glissant ses mains sous l’élastique de sa culotte de dentelles, il se mit à lui pétrir les fesses avant de venir taquiner son intimité.*

Le stéréotype sexuel — l’homme entreprenant — est encore à l’œuvre ici.

Le contact est réitéré :

il le V-PS de nouveau (inf. ; 59/59)

(11) *Il l’embrassa de nouveau, plus passionnément cette fois.*

(12) *Il la posséda de nouveau avec une force tendrement sauvage.*

Le contact est l’aboutissement d’un mouvement impliquant le corps entier de l’actant sujet :

il se V-PS contre lui¹¹ (inf. ; 59/59)

(13) *Sans un mot, elle se lova contre lui et ne bougea plus.*

(14) *Haletante, elle se cambra contre lui.*

(15) *Elle se plaqua contre lui, lui faisant sentir la force de son désir.*

Le personnage féminin est toujours à l’initiative du mouvement dont le terme est l’actant masculin constitué comme force statique. Ce mou-

11. Ne pas oublier que **il**, **lui** dans la notation des motifs, sont des lemmes qui valent donc pour *il*, *ils*, *elle*, *elles*.

vement constitue une phase préparatoire¹², avant le début de la relation charnelle ; si le personnage masculin, comme on l'a vu avec le motif **il lui V-PS le NC**, est à l'initiative de cette relation (sa phase inchoative), sa préparation est de la responsabilité de l'héroïne.

3.1.2. La relation entre les deux actants est un mouvement ou un geste orienté

L'actant objet est destinataire d'une manifestation physique de l'actant sujet :

il lui V-PS un ADJ NC (inf. ; 51/51)

(16) *Il lui adressa un faible sourire, évitant de croiser son regard.*

(17) *Il lui fit un petit signe de tête.*

(18) *Elle lui donna un gros baiser sonore.*

L'adjectif qualifiant l'objet communiqué (*faible, petit, gros*) est révélateur de l'intensité de la relation entre les deux actants.

Il y a mouvement vers l'actant objet ; ce mouvement constitue le préalable à un contact. Syntaxiquement, il y a donc deux phrases coordonnées :

il se V-PS vers lui et V-PS (inf. ; 51/51)

(19) *Il se tourna vers elle et surprit son regard posé sur lui.*

(20) *Il s'avança vers elle et déposa un baiser sur ses lèvres.*

(21) *Il se pencha vers elle et prit ses lèvres chaudes et sensuelles.*

3.1.3. La relation entre les deux actants est de nature perceptuelle

La relation entre l'actant sujet et l'actant objet se manifeste par le regard (plus rarement par l'ouïe) ; le regard est caractérisé par une réaction physique de l'actant sujet :

il le V-PS de un NC ADJ (inf. ; 52/52)

(22) *Elle le gratifia d'une œillade sévère.*

(23) *Elle l'enveloppa d'un regard furieux, accusateur.*

(24) *Elle l'écouta d'une oreille distraite.*

12. La terminologie aspectuelle manque d'érotisme. La phase préparatoire équivaut sans doute aux préliminaires.

Air sidéré, regard sombre, œil sceptique, etc. : la réaction physique est très systématiquement négative. On compte très peu de *regard caressant, air tranquille*.

3.2. Le motif exprime le rapport de l'actant sujet avec ses émotions / passions / sentiments

3.2.1. L'actant sujet parvient à maîtriser ses émotions / passions / sentiments

La sollicitation de l'émotion est réprimée :

il se **V-PS** de **INF** le **NC** (inf. ; 80/91)

(25) *Elle s'efforça de refouler les souvenirs de leur vie commune.*

(26) *Elle s'empessa de détourner la tête.*

La tentation est à son paroxysme :

<**devoir se faire violence pour ne pas**> (inf. ; 16/16)

(27) *Ce simple contact la fit fondre, et elle dut se faire violence pour ne pas céder au désir.*

(28) *Carter dut se faire violence pour ne pas goûter séance tenante au nectar de ses lèvres adorables.*

3.2.2. Le sujet actant ne maîtrise pas la passion

La défaillance du contrôle de la passion est cause de déception :

à **DETPOSS** **ADJ** **NC**, (inf. ; 93/117)

(29) *À sa grande honte, elle avait besoin de lui.*

(30) *À sa grande consternation, elle commençait à s'apercevoir que cet homme l'obsédait.*

Le contrôle de la passion est une mise à l'épreuve de l'actant sujet :

ADVE **ADJ** **que** **il** **V-PS** (inf. ; 36/36)

(31) *Alexius sentait Rosie si excitée qu'il dut lutter pour se contenir.*

(32) *Ce son lui parut tellement définitif qu'elle sentit ses genoux fléchir.*

Le motif prend ici la forme d'une proposition consécutive intensive¹³.

13. Voir Adam (2011) pour une étude de cette construction dans les genres de l'insulte rituelle, du conte (de Perrault) et de l'argumentation publicitaire. Cf. aussi Plantin (1985).

L'émotion n'est pas contrôlable, car elle est invasive :

un NC de NC le V-PS (inf. ; 105/130)

(33) *Une vague de chaleur la submergea quand elle le revit ramasser ses jetons sur le tapis vert.*

(34) *Un plaisir intense et un sentiment de bien-être la gagnèrent.*

(35) *Un frisson d'excitation la parcourut.*

La passion est une force irrésistible :

<mais ce être plus fort que il> (inf. ; 10/10)

(36) *Non, ce n'était pas juste, ni même raisonnable de réagir ainsi, mais c'était plus fort qu'elle.*

(37) *Elle savait qu'elle le dévorait du regard de la manière la moins discrète qui soit, mais c'était plus fort qu'elle.*

C'est la relation modale entre le «savoir se comporter» (savoir portant en fait sur un système social déontique : le convenable, le raisonnable) et le «pouvoir se comporter» qui est articulé dans cette routine lexicalisée.

La force de l'émotion, du sentiment, est inédite :

<jamais il ne se être sentir aussi ADJ> (inf. ; 11/11)

(38) *Jamais elle ne s'était sentie aussi désespérée, aussi perdue et seule.*

(39) *Jamais il n'avait éprouvé une telle impression de communion avec un être.*

La force de l'émotion, du sentiment est trop forte :

<je ne pouvoir pas me INF de> (inf. ; 16/16)

(40) *Au restaurant, je ne pouvais pas m'empêcher de te regarder.*

(41) *Irina, je ne peux pas me permettre de vous perdre, déclara-t-il.*

<se laisser INF par le NC de> (inf. ; 20/20)

(42) *Haletante, frémissante, elle se laissa emporter par les spasmes de l'orgasme.*

(43) *Elle avait déjà bien assez de mal à ne pas se laisser distraire par la pensée de Justin.*

Ici, le sujet est toujours féminin. Les deux derniers motifs mettent en évidence la relation naturelle entre une sémantique modale (*ne pas pouvoir / se laisser*) et une sémantique des passions.

3.3. Le motif exprime une manifestation physique de l'émotion

3.3.1. L'émotion est visible

L'émotion se donne à voir sur le visage du personnage :

le NC de NP se V-PS de NC (inf. ; 42/42)

(44) *Le visage de Tessa s'éclaira de joie.*

(45) *Les yeux de Roth s'arrondirent de surprise.*

Le point de vue de la perception n'est pas situé (il n'a pas pour origine un personnage). Ce qui n'est pas le cas pour le motif suivant où :

L'actant sujet est témoin d'une transformation physiologique de l'actant objet causée par l'émotion :

voir le NC de NP se INF (inf. ; 36/36)

(46) *Elle vit les yeux de Luz s'écarter de surprise.*

(47) *Eliza fut surprise de voir le visage de Mary se rembrunir.*

(48) *Elle vit la mâchoire de Luke se contracter.*

La subordonnée infinitive est ici mobilisée.

3.3.2. L'émotion ou le sentiment est cause d'une réaction physique du sujet actant

L'émotion se manifeste par des effets physiologiques que perçoit par proprioception l'actant sujet :

NP sentir DETPOSS NC se INF (inf. ; 93/104)

(49) *Kara sentit son cœur se serrer tandis que Declan caressait les cheveux de Safia.*

(50) *Katrina sentit ses joues s'empourprer.*

(51) *Natalie sentit ses épaules s'affaïsser.*

L'émotion se manifeste par des effets psychologiques ; s'ensuit une réaction corporelle :

PASS¹⁴, il V-PS le NC (inf. ; 81/107)

(52) *Choquée, elle hocha la tête, tandis que Greg sortait son téléphone portable de sa poche.*

(53) *Agacée, elle leva les yeux au ciel et ne bougea pas d'un pouce.*

(54) *Accablée, elle haussa les épaules.*

14. PASS = participe-passé.

Il s'agit là d'un motif exclusivement «féminin», prenant la forme d'une construction détachée. Le participe passé en position frontale a presque toujours une valeur dysphorique : les mouvements du corps, ritualisés, marquent bien souvent la consternation, le désappointement, etc.

3.4. *Changement dans la fabula*

Il existe bien sûr, dans ces histoires stéréotypées, un moment où une situation se coupe du passé. L'expression *la première fois*, statistiquement sur-employée dans le sous-genre, joue le rôle narratif, dans des configurations syntaxiques différentes, d'exprimer une époque révolue, et l'avènement d'une nouvelle situation importante dans l'histoire :

<pour la première fois depuis que il>

(55) *Pour la première fois depuis qu'ils s'étaient retrouvés, elle prit conscience de ce qui avait changé en lui.*

<pour la première fois depuis DETPOSS NC> (inf.; 14/14)

(56) *Pour la première fois depuis son adolescence, il se surprenait à s'amuser comme un gosse.*

<ce être la première fois que je> (inf.; 17/17), **<ce être la première fois que NP>** (inf.; 11/11)

(57) *Je te l'ai dit, c'est la première fois que je fais une telle déclaration.*

(58) *C'était la première fois que Tamara s'autorisait à exprimer ses craintes tout haut.*

Le segment suivant remplit une fonction semblable :

<Il ne y avoir pas si longtemps> (inf.; 11/11)

(59) *Il n'y avait pas si longtemps, Travis était convaincu que Jericho ne s'assagirait jamais.*

3.5. *Autres motifs pourvoyeurs de clichés*

Terminons avec quelques exemples — bien d'autres pourraient être donnés — de clichés spécifiques du genre sentimental :

L'actant sujet initie un geste ou un comportement préalable à l'accomplissement d'un procès :

il V-PS le NC_{corps} pour INF (inf.; 81/105)

(60) *Il leva le bras pour saisir une branche de pommier qui pendait au-dessus de sa tête.*

(61) *Elle baissa la tête pour observer ses escarpins.*

(62) *Il tendit le cou pour regarder l'arrière du véhicule.*

Il y a un rapport quasi pléonastique entre les deux procès, dans la mesure où le second présuppose le premier, qui est informativement non pertinent, mais permet à l'auteur de remplir des pages de texte.

L'attitude du personnage est confortée, renforcée, confirmée par un facteur :

ne faire que INF (inf. ; 45/45)

(63) *L'aveu de mon père ne fait que renforcer mes certitudes.*

(64) *La colère contenue de la jeune femme ne fit qu'amuser un peu plus Bill.*

(65) *Ce silence ne fit qu'accentuer le trouble de Sophie.*

3.6. Motifs non clichés mais structurants

Enfin, il existe des motifs qui ne sont pas à proprement parler des clichés, puisque leur valeur est plutôt fonctionnelle, mais qui néanmoins se caractérisent par leur récurrence : ils marquent, tout autant que les autres, la signature du genre. Ainsi en est-il des motifs dédiés à la construction de la relation premier plan/arrière-plan dans la narration. Trois motifs assurent cette fonction ; ils prennent la forme d'une subordonnée circonstancielle à valeur temporelle, employée en début de phrase ; classiquement, le verbe de la subordonnée est à l'imparfait ou au passé antérieur¹⁵ :

comme il VIMP le NC (inf. ; 56/56)

(66) *Comme il reposait le cadre à sa place, elle se leva pour le rejoindre.*

(67) *Comme elle regagnait le bureau central, elle y retrouva Sam.*

tandis que il se VIMP (inf. ; 124/166)

(68) *Tandis qu'elle s'installait à un lavabo à côté de lui, elle le vit sourire sous son masque.*

(69) *Tandis qu'ils se souriaient, une épaisse mèche de cheveux agitée par la brise effleura son visage.*

15. Précisons : ce n'est pas la fonction qui est spécifique au genre, mais l'emploi statistiquement significatif des mêmes motifs pour remplir cette fonction.

dès que il avoir PASS le NC (inf. : 70/88)

(70) *Dès qu'il eut ouvert la porte d'entrée, Callie ne put cacher sa surprise.*

(71) *Dès qu'il eut quitté la pièce, le masque tomba.*

La phraséologie considérée ici ne caractérise pas tant la structure interne de la forme, que sa relation avec une pratique d'écriture. Il s'agit d'une manière caractéristique de dire, de composer le texte sentimental, pouvant s'instancier dans des propositions formellement différentes, mais partageant cependant des traits invariants. Autrement dit, l'usage de la forme est ici autant phraséologique que la forme elle-même.

Ces quelques exemples de clichés renforcent une conception que nous défendons : la nature du cliché n'est pas à chercher seulement dans les formes mêmes. Si le cliché était une pure question de forme, rien ne le distinguerait des autres unités phraséologiques. Les clichés ne sont, en quelque sorte, que des indices d'un procédé. Ce qui *fait cliché*, c'est la reconnaissance *dans* ces formes d'un procédé d'écriture : le cliché est dans la mise en pratique (d'écriture) systématique d'une intention esthétique et stylistique. Par exemple **il lui V-PS le NC** (*il lui prit les lèvres*) n'est pas un cliché parce qu'il instancie des énoncés dans lesquels figure un champ lexical marqué (verbe de contact, partie du corps). Il l'est parce que la thématique qu'il traduit (le contact sensuel entre un homme et une femme, dans un contexte de romance), loin de pouvoir être exprimée par un ensemble non fini de formes (comme dans d'autres genres) est, ici, dépendante de quelques formes seulement. Autrement dit, il y a cliché lorsqu'il y a restriction du champ expressif d'une thématique. Ailleurs, dans un autre contexte d'emploi, le même motif, les mêmes phrases, n'auraient pas nécessairement statut de clichés.

Conclusion

Nous avons illustré la manière dont on peut, grâce aux motifs, travailler la notion de cliché : non pas partir d'une liste préconstruite de clichés, mais plutôt d'une exploration des textes et de leurs formes, pour qu'émergent de façon non (ou peu) supervisée des unités à partir desquelles l'interprétation peut être conduite.

La méthode que nous avons présentée ici n'est pas seulement un moyen efficace pour identifier des patrons flexibles ; permettant d'extraire des unités qui ne sont pas décelables (ou facilement décelables) par les méthodes traditionnelles, elle met en évidence un ensemble élargi

de formes — ce qui constitue la condition nécessaire pour l'analyse du pré-construit dans les genres. Aussi cette méthode contribue-t-elle également à reconsidérer les études sur la phraséologie, en proposant une « extension du domaine » (Legallois & Tutin, 2013). Il n'en demeure pas moins, évidemment, qu'elle reste encore largement à améliorer. Plusieurs pistes sont envisagées dans des futurs travaux, comme :

- la prise en compte de variabilité des motifs, par exemple leur propension à connaître des expansions ou des réductions ;
- la co-occurrence de motifs : quels sont les motifs qui ont tendance à être employés dans le même contexte (phrases, paragraphes, chapitres) ;
- la topologie : où, dans le texte, tel motif a-t-il tendance à apparaître (colligation textuelle) ?

Du pain sur la planche... (pour finir sur un cliché).

Remerciements

Ce travail a bénéficié d'une aide de l'État gérée par l'Agence nationale de la recherche au titre du programme « Investissements d'avenir » portant la référence ANR-10-LABX-0083.

RÉFÉRENCES BIBLIOGRAPHIQUES

- ADAM, Jean-Michel. (2011). Les consécutives intensives : un schéma syntaxique commun à plusieurs genres de discours. *Linx*, 64-65, 115-131. Disponible en ligne sur <<http://linx.revues.org/1407>>.
- AMOSSY, Ruth & ROSEN, Elisheva. (1982). *Les discours du cliché*. Paris : Éditions CDU et SEDES.
- AMOSSY, Ruth & HERSCHBERG PIERROT, Anne. (1997). *Stéréotypes et clichés. Langue, discours, société*. Paris : Nathan Université.
- BARLOW, Michael. (2004). *Collocate 1.0. Locating Collocation and Terminology*. Houston, TX : Althelstan.
- BEAUVISAGE, Thomas. (2001). Exploiter des données morphosyntaxiques pour l'étude statistique des genres. Application au roman policier. *Traitement automatique des langues*, 43, 45-65.
- BIBER, Douglas. (1988). *Variation across Speech and Writing*. Cambridge : Cambridge University Press.
- CHURCH, Kenneth & HANKS, Patrick. (1990). Word Association Norms, Mutual Information, and Lexicography. *Computational Linguistics*, 16, 22-29.

- CONSTANS, Ellen. (1999). *Parlez-moi d'Amour. Le roman sentimental, des romans grecs aux collections de l'an 2000*. Limoges : Presses universitaires de Limoges.
- DIWERSY, Sascha & LEGALLOIS, Dominique. (À paraître). L'apport de la méthode des motifs aux analyses phraséologiques en Discours. Dans *Actes du Colloque international « Approches théoriques et empiriques en phraséologie »*. Tübingen : Stauffenburg.
- DUFAYS, Jean-Louis. (1994). *Stéréotype et lecture*. Liège : Mardaga.
- ECO, Umberto. (1979/1985). *Lector in Fabula*. Paris : Grasset.
- HOUEL, Annik. (1997). *Le roman d'amour et sa lectrice. Une si longue passion*. Paris : L'Harmattan.
- LEGALLOIS, Dominique. (À paraître). *Caractériser le style d'un auteur par des patrons lexico-grammaticaux : une nouvelle approche en stylistique*. Communication présentée au Colloque international « Méthodes stylistiques. Unités et paliers de pertinence ? », Lyon.
- LEGALLOIS, Dominique & TUTIN, Agnès. (2013). Présentation : Vers une extension du domaine de la phraséologie. *Langages*, 189, 3-25.
- LEGALLOIS, Dominique & PRUNET, Anne. (2015). Sequential Patterns: A New Corpus-Based Method to Inform the Teaching of Language for Specific Purposes. *Journal of Social Science*, 44, 127-140.
- LONGRÉE, Dominique & MELLET, Sylvie. (2013). Le motif : une unité phraséologique englobante ? Étendre le champ de la phraséologie de la langue au discours. *Langages*, 189, 68-80. <<http://dx.doi.org/10.3917/lang.189.0065>>.
- MALRIEU, Denise & RASTIER, François. (2001). Genres et variations morphosyntaxiques. *Traitement automatique des langues*, 42(2), 548-577.
- MARION, François. (2009). Le stéréotype dans le roman policier. *Cahiers de Narratologie*, 17. Disponible en ligne sur <<http://narratologie.revues.org/1095>> (consulté le 3 juillet 2015).
- MEL'ČUK, Igor. (1996). Lexical Functions: A Tool for the Description of Lexical Relations in the Lexicon. Dans L. Wanner (dir.), *Lexical Functions in lexicography and Natural Language Processing* (p. 37-102). Amsterdam/Philadelphie : John Benjamins.
- PÉQUIGNOT, Bruno. (1991). *La relation amoureuse. Analyse sociologique du roman sentimental moderne*. Paris : L'Harmattan.
- PLANTIN, Christian. (1985). La genèse discursive de l'intensité : le cas du SI « intensif ». *Langages*, 80, 35-53.
- QUINIOU, Solen, CELLIER, Peggy, CHARNOIS, Thierry & LEGALLOIS, Dominique. (2012). What about Sequential Data Mining Techniques to Identify Linguistic Patterns for Stylistics? Dans A. Gelbukh (dir.), *Computational Linguistics and Intelligent Text Processing* (vol. 7181, p. 166-177). Berlin/Heidelberg : Springer.

- RICHAUDEAU, François. (1986). La galaxie Harlequin, des auteurs et des romans. *Communication et langages*, 67, 9-24.
- RIFFATERRE, Michael (1964). Fonctions du cliché dans la prose littéraire. *Cahiers de l'Association internationale des études françaises*, 16, 81-95.
- SALEM, André. (1987). *Pratique des segments répétés. Essai de statistique textuelle*. Paris : Klincksieck.
- SUEUR, Jean-Pierre. (1982). Pour une grammaire du discours. Élaboration d'une méthode ; exemples d'application. *Mots*, 5(1), 143-185.